A New Method to Quantify the Causal Effects of Reinforcement in Terms of Behavioral Selection

Matthias Borgstede, Patrick Anselme***

*Corresponding author:

University of Bamberg

Markusplatz 3

D-96047 Bamberg

matthias.borgstede@uni-bamberg.de

**Ruhr-University Bochum

D-44780 Bochum, Germany

patrick.anselme@rub.de

A New Method to Quantify the Causal Effects of Reinforcement in Terms of Behavioral Selection

Abstract

We present a new methodology to partition different sources of learning within a selectionist framework based on the Price equation – the Multilevel Model of Behavioral Selection (MLBS). The MLBS provides a theoretical and formal background to disentangle behavior changes due to selection effects ("reinforcement") from behavior changes due to non-selection effects (e.g., "effort" or "skill"). The model allows to quantify the causal effect of reinforcement even in the presence of opposing sources of behavior change and to test whether the reinforcement effect is statistically significant for a given subject.

We demonstrate our method using training data from a reinforcement experiment with nine pigeons (*Columba livia*). The pigeons were trained to search for food in small holes on the ground of a special cage. In each session, the holes in only one half of the cage were filled with food (reinforced region), while the holes in the other half of the cage remained empty (nonreinforced region). Although the total time spent in the reinforced region between training sessions tends to decrease between training sessions, our method reveals positive behavioral selection (reinforcement effect) in most of the pigeons. However, this selection effect is masked by a strong non-selection component that consists in pigeons becoming more efficient foragers (skill effect) and in an increased peck frequency (effort effect).

The partitioning of behavior change into selection and non-selection components allows to identify the causal effects of reinforcement even if they cannot be isolated experimentally. Our method is especially useful in the context of non-equilibrium behavior analysis (such as the change between training sessions), and in the context of behavior observations in naturalistic settings.

Keywords: *reinforcement, behavioral selection, Price equation, covariance based law of effect, multilevel model of behavioral selection, causal inference*

1 Introduction

The quantitative analysis of behavior traditionally focuses on establishing functional relations between behavior and the contextual factors in which the observed behavior occurs, such that one or more behavioral measures (e.g., time spent doing an activity, or lever press rate) is expressed as a mathematical function of one or more independent variables (e.g., delay of reinforcement or the amount of reinforcement). For example, in matching analysis, the ratio between two observed behavioral measures, such as the time spent engaging in two mutually exclusive activities (e.g., pressing one of two different levers), is described as a function of the ratio between the amount of reinforcement received during the different activities (e.g., food pellets) (Herrnstein, 1961). Functional relations may help structure observed data, predict future behavior and advance new ways to influence behavior. Furthermore, given all other factors are held constant in an experiment, a functional relation between an independent and a dependent variable allows us to infer a causal effect of the independent variable (Menzies & Price, 1993). The experimental method has revealed many consistent qualitative findings in the field of behavior analysis, such as the causal effects of different schedules of reinforcement (Ferster & Skinner, 1957), exposure to stressful stimuli (Lemos et al., 2012), and drug consumption (Berridge & Robinson, 2016), on behavior change.

However, when reinforcement is studied on a quantitative level, causal inference is not quite as straightforward. The problem with reinforcement is that, in order to be effective, it needs to be contingent on behavior. Therefore, the received amount of reinforcement in a behavioral experiment does not only depend on the experimenter, but also on the behavior of the individual under study. Consequently, it is impossible to isolate reinforcement as an independent variable. For example, as pointed out by Staddon (2020), the matching law is a functional relation between two *dependent* variables – a behavior ratio and a reinforcement ratio. Therefore, one cannot interpret the

parameters of the fitted matching equation in terms of causal effects of reinforcement. Since reinforcement always depends on behavior, this problem persists even if the data come from the most controlled experiments.

The situation becomes even more difficult when behavior is studied under non-equilibrium conditions, such as when the dynamics of behavior change over time due to gradual depletion in reinforcers from trial to trial. As long as behavior changes dynamically, the experimenter has little control over the amount of received reinforcement. In fact, conditional on the schedule of reinforcement, the amount of received reinforcement solely depends on the behavior of the individual under study, making it impossible to isolate the causal effect of reinforcement experimentally.

Although the above methodological problems are well known, they are hardly ever considered in the analysis of behavioral experiments. From a pragmatic point of view, focusing on functional, rather than causal, relations may seem reasonable. However, even if one is "only" interested in the prediction and control of behavior, it is difficult to adhere to a purely functional account. In fact, the very idea of controlling behavior implies that the experimentally identified determinants of behavior must be effective *causes* of behavior. Moreover, if there is one thing that behavior analysts can agree upon, it is certainly that behavior changes *because* of reinforcement.

But in which sense can reinforcement be said to be a cause of behavior? Following Skinner's three-term contingency, reinforcement requires the occurrence of an event (such as the administration of food) contingent on a specific behavior (such as lever pressing), given specific contextual stimuli (such as a flashing light in an operant chamber) (Skinner, 1965). If we observe a quantitative behavior change in the presence of the contextual stimuli, we call the contingent event a reinforcer and attribute the observed change in behavior to reinforcement. Applied to experimental data, the three-term contingency yields an operational definition of reinforcement. All other things being equal, we could say that behavior changed *because* of reinforcement, since all other potential causes of

behavior change were held constant and could therefore be ruled out. In other words, if we want to give a causal interpretation to reinforcement, we need to treat it as an independent variable and hold all confounding variables constant. However, as argued above, it is impossible to isolate the amount of reinforcement as an independent variable in an experiment because reinforcement is necessarily contingent on behavior. Therefore, although the three-term contingency provides a coherent conceptual framework for the analysis of behavior, is not sufficient to justify a causal interpretation of reinforcement.

To give an example, imagine a pigeon searching for food in an unfamiliar environment. We may quantify its foraging behavior by measuring the time spent at various food patches over several successive time intervals. Say that one food patch contains significantly more food than the other patches. If reinforcement was effective, we would expect the pigeon to eventually spend more time in the food patch that contains more food. However, other factors may also be effective in altering behavior. For example, the pigeon may follow a very simple, fixed decision rule like this one: "As long as you keep finding food, stay close; else, move to another location" (cf. McNamara & Houston, 2009). Such a rule does not require any kind of reinforcement or long-term learning. However, it will naturally produce a quantitative change in behavior over successive observation intervals, such that the pigeon will end up spending most of its foraging time in the region with the largest amount of food. Accordingly, pigeons prefer to forage in patches with a higher food density, especially if the food items can be detected visually (Anselme et al., 2018; Anselme et al., in press). Even if all external factors were held constant, we could not be sure whether the observed change was actually caused by reinforcement, or by a simple local decision rule that mimics reinforcement but is in fact something completely different. Reinforcement could be effective but remain undetectable because other sources of behavior change may mask its causal effect on behavior. For example, the pigeon may not only learn where the food is, but also increase its foraging effort by means of a higher pecking rate. It may also acquire skills that facilitate food retrieval over time. Both factors will result in the pigeon spending less time in the rich food patch because they enable the pigeon to be a more

efficient forager. If the effects of effort and skill are strong enough, they may actually mask the effect of reinforcement or even result in a "negative" overall behavior change—i.e., a lower investment in the task. If we attributed this negative change to reinforcement effects, we would have to conclude that the obtained food acts as a punisher, due to the pigeon eventually spending less time at the rich food patch.

The above example illustrates that the causal effect of reinforcement cannot easily be inferred from observed behavior change. Other sources of behavior change may mimic reinforcement, even if the individual is never actually reinforced, or mask existing reinforcement effects. Such non-reinforcement factors may be difficult to control because: a) they may be unobservable, such as local decision rules, b) they may be dynamic in nature and thus change in the course of the experiment, such as skills, or c) they lie outside the influence of the experimenter, such as increased effort. Since reinforcement itself cannot be isolated experimentally, non-reinforcement sources of behavior change may occur in the most controlled experiments, and will probably be even more important in the wild. Indeed, some findings indicate that animals often spend time and effort seeking uncertain food despite the availability of large amounts of free food (Inglis et al., 1997), accept to pay a cost for an information about a future outcome that is independent of this request for information (Wang & Hayden, 2019), and sometimes develop various pathologies in captivity when the opportunity to exert the actions leading to reward procurement in nature is not possible, despite the absence of food deprivation (Eilam et al., 2006). These phenomena suggest that reinforcement is not all in animal behavior.

In this paper, we address the issue of causal inference in the context of reinforcement. First, we provide a coherent theoretical background for reinforcement that builds on the concept of selection by consequences (Skinner, 1981). Second, we use the selectionist theory of reinforcement to develop a new methodology that allows for the quantification of the causal effects of reinforcement by means of counterfactual reasoning (Lewis, 1973). Third, we apply our methodology to quantify the

causal effects of reinforcement in the presence of other sources of behavior change using training data from a pigeon experiment.

The remainder of this article starts with an introduction to the selectionist account of reinforcement, and its formalization within the Multilevel Model of Behavioral Selection (MLBS), in particular (Borgstede & Eggert, 2021). We then apply the rationale of counterfactual reasoning to derive a new method for the quantification of the causal effects of reinforcement on behavior change between experimental trials. The method also allows for the calculation of p-values to test the selection effect for statistical significance on the level of single subjects, which can then be aggregated to the group-level using a meta-analytic approach (cf. Dugard et al., 2012). Finally, we apply our method to an empirical data set, showing that reinforcement may be causally effective even if it is masked by other factors, and discuss the implications with regard to the selectionist theory of reinforcement and potential practical applications of the method.

2 Behavioral Selection Theory

Reinforcement has repeatedly been characterized as a process that works analogous to natural selection (Broadbent, 1961; D. T. Campbell, 1956; Gilbert, 1970; Herrnstein, 1964; Palmer & Donahoe, 1992; Pringle, 1951; Skinner, 1966; Thorndike, 1900). Whereas in natural selection, species adapt to the environment as a result of the fitness consequences of inheritable traits, reinforcement consists of individuals adapting to specific contexts as a result of the reinforcing consequences of repeatable actions. Skinner (1981) proposed that both processes should be subsumed under the common explanatory mode of “selection by consequences.”

The idea that reinforcement may be a selection process was later formalized by Baum (2017) and further developed by Borgstede and Eggert (2021), who integrated individual level behavioral selection with population level natural selection within a Multilevel Model of Behavioral Selection (MLBS). The MLBS builds on the abstract description of selection processes by means of the Price equation (Price, 1970, 1972). The Price equation describes selection by means of the covariance

between the individual values of a quantitative character (e.g., an inheritable trait such as size) and individual evolutionary fitness (i.e., the contribution of an individual to the future population). A positive covariance is associated with a positive change in mean character value (i.e., selection results in a higher average value), whereas a negative covariance indicates a negative change (i.e., selection results in a lower average value). In other words, those character values that coincide with higher evolutionary fitness, are selected and thus alter the population average. The Price equation provides a mathematically exact description of this process that is independent of the specific mechanisms of variation, selection, and transmission (Luque, 2017).

In the Price equation, all other sources of change (those that do not refer to selection) are subsumed in a residual term. This term has different interpretations depending on the context. For example, when applied to the evolutionary change of gene frequencies, the residual term captures the effects of imperfect transmission (i.e., mutation and recombination). When applied to phenotypic change over generations (e.g., body weight or behavioral traits), the term may also capture environmental factors influencing the phenotype (Luque & Baravalle, 2021).

In the MLBS, the Price equation framework is applied to a population of individuals that vary in a certain quantitative behavior (such as average time spent doing a certain activity). In this case, the covariance term refers to natural selection on average behavior tendencies of the population, and the residual term captures the average change within individuals. The MLBS is mainly concerned with this latter within-individuals change. Following the rationale that individual changes in behavior can be explained through selection by consequences, the MLBS extends the Price equation by applying the same covariance principle on the within-individuals level. Here, the population average in behavior is not calculated over different individuals, but over recurring instances of the same context (e.g., experimental trials). Behavior change, such as an increase or decrease in time spent doing a certain activity averaged over a longer period, is described according to the covariance partitioning from the Price equation. However, the criterion of selection is not individual fitness itself (in terms of a direct contribution to the future population), but statistical predictors of individual fitness. In other

words, on the within-individual level, behavior is not selected by means of reproduction or survival (in fact, if the individual dies, all of its behavior immediately ceases), but by events that predict expected change in evolutionary fitness (Borgstede, 2020). For example, food is generally a positive predictor of evolutionary fitness because it raises the probability of survival and, thus, future reproduction. Conversely, physical threat is a negative predictor of evolutionary fitness because it lowers the chances of survival and future reproduction. The concept of a fitness predictor is largely equivalent to what Baum (2012) calls a "phylogenetically important event (PIE)". The only difference is that, in the MLBS, fitness predictors are defined on purely statistical grounds, whereas Baum emphasizes the causal effects of PIEs on evolutionary fitness.

The conceptual framework of the MLBS allows us to describe behavioral selection on the individual level by means of a within-individuals covariance between behavior in recurring contexts and its consequences in terms of statistical fitness predictors. A positive covariance corresponds to an increase in behavior due to reinforcement, whereas a negative covariance implies a decrease in behavior due to punishment (cf. Baum, 2012). Similar to the process of natural selection, those behaviors that coincide with events that signal a gain in expected evolutionary fitness are selected, which in turn changes the average behavior of the individual. Following the MLBS, the amount of change due to selection corresponds to the covariance between behavior and fitness predictor, weighted by the slope of the fitness predictor (its "reinforcing power"). Formally, the amount of behavior change due to within-individual selection can be expressed by the following equation (Borgstede & Eggert, 2021):

$$\Delta_S b = \beta_{wp} \operatorname{Cov}(b,p) \tag{1.}$$

Here, b designates a quantitative behavior (such as time spent doing a certain activity) and $\Delta_S b$ corresponds to the within-individual change in average behavior due to selection. The value p is a quantitative fitness predictor and $\operatorname{Cov}(b,p)$ is the covariance between the behavior and the fitness

predictor. Finally, β_{wp} is the slope of the fitness function of the predictor p (i.e., its reinforcing power).

Just like the original Price equation, the MLBS introduces a residual term that captures all other sources of behavior change that are not selection. Designating the overall change in average behavior, Δb is the within-individuals behavior change which can now be expressed as the sum of a selection term and a residual term δ:

$$\Delta b = \Delta_S b + \delta \tag{2.}$$

This equation is called the covariance based law of effect (CLOE). The CLOE can be regarded as a fundamental principle of behavior analysis in that it captures the essence of behavioral selection by partitioning the overall change in behavior into a selection component and a residual component (Borgstede & Luque, 2021). The selection component corresponds to what is traditionally called "reinforcement", whereas the residual term subsumes all other sources of behavior change that are not reinforcement. In the following section, we use the MLBS as a starting point for a theory-based approach to measuring the effects of reinforcement (cf. Borgstede & Eggert, in press).

3 Causal Analysis and Counterfactual Reasoning

On a theoretical level, the MLBS accounts for a great range of behavioral phenomena that are difficult to explain using traditional stimulus-response theories (cf. Borgstede, 2021; Borgstede & Eggert, 2021; Borgstede & Luque, 2021), such as the blocking effect (Kamin, 1969), reinforcing effects of rare behavior (Premack & Premack, 1963), reinforcing effects of response deprivation (Timberlake & Allison, 1974), information seeking (Berlyne, 1957; Hendry, 1965), and various deviations from matching (Davison & McCarthy, 2016).

However, quantitative estimates of behavioral selection from empirical data can be difficult to obtain (cf. Strand et al., 2021). As mentioned before, selection effects may be masked by other sources of behavior change, which is why a direct correspondence between observed behavior change and

behavioral selection is not guaranteed. Moreover, observed behavior may differ between trials due to random fluctuations, rather than selection. Therefore, estimates of behavioral selection from empirical covariances between behavior and reinforcement are only reliable for large amounts of data. However, as long as selection is present, behavior is not in an equilibrium state and thus is subject to ongoing change, which in turn biases covariance estimates.

The above challenges are not unique to the theory of behavioral selection, but also occur in biological applications of the Price equation. Although the partitioning of change into selection and non-selection components is always possible on a theoretical level, empirical applications require specific models about the causes of change (Okasha & Otsuka, 2020). Okasha and Otsuka (2020) propose to use such models for causal inference using counterfactual reasoning (Lewis, 1973, 2000). Given a specific model, counterfactual analysis yields empirical estimates of the causal effects of the selection and non-selection components in the Price equation. In the domain of behavioral selection, a causal interpretation of the selection term in the CLOE would provide a quantitative estimate of the causal effect of reinforcement. Therefore, a method based on counterfactual reasoning might provide a tractable method to infer the causal effects of reinforcement from observed behavioral change.

The general idea behind counterfactual reasoning is that we attribute a causal effect to an event if we have reason to believe that, all other things being equal, the effect would not have occurred in the absence of the event (Pearl, 2013). Given this interpretation of causality, conducting a controlled experiment is an attempt to produce two exchangeable scenarios that are identical in every relevant respect except for the event that we assume to be the cause of what we expect to observe. In between-subject experiments, we may ensure exchangeability by randomly assigning the individuals to the experimental conditions. In within-subject experiments, we may ensure exchangeability by randomly choosing the temporal order of the experimental condition for each subject (Edgington & Onghena, 2007). Applying counterfactual reasoning, a properly designed experiment shows us, what

would have happened to an individual in condition A, if it had been assigned to condition B, instead (and vice versa).

As mentioned in section 1, it is sometimes difficult or even impossible to realize experimental conditions that are equal in every relevant respect except for the independent variable. Nevertheless, we may still apply the method of counterfactual reasoning to justify causal inference, as long as we make our inference conditional on a specific model. In other words, given a theoretical model, we may use empirical observations to estimate what would have happened if certain parameters had been equal. In this sense we may justify a claim like "the gravitational field of the sun causes the earth to move on its observed trajectory", although it is technically, if not physically, impossible to change the gravitational field of the sun in a controlled experiment to verify our causal claim. We may, however, justify this causal claim based on a theoretical model that describes the trajectory of planets in different gravitational fields (e.g., classical mechanics). We can apply the model to empirical observations regarding the movement of the earth and the other planets in our solar system and use these data to estimate the gravitational pull of the sun on the earth, as well as the earth's momentum. Together, these empirical estimates yield a very accurate description of the earth's trajectory. A causal interpretation can now be justified by leaving the parameterized model unchanged, except for the parameters that corresponds to the gravitational field of the sun. If we set the sun's gravitation equal to zero (in our model!), we may then calculate what would have happened if the sun had not kept the earth in orbit with its gravitational field. The difference of these two trajectories – the actually observed one and the counterfactual one – quantifies the causal effect of the sun's gravitational field.

Similar to the above physical example, we can justify theory-based causal inference using counterfactual reasoning in behavior analysis. Such a theory-based approach would be especially helpful in situations where it is difficult or impossible to justify a causal interpretation on experimental grounds. Therefore, here we propose theory-based approach to infer the causal effects

of reinforcement from observed behavior change by means of the above described Multilevel Model of Behavioral Selection (MLBS) and the Covariance Based Law of Effect (CLOE).

The CLOE describes change due to reinforcement by means of a covariance term, such that a higher covariance between behavior and reinforcement yields a higher selection and a lower covariance yields smaller selection. To apply the counterfactual method outlined above, we need to use the MLBS in the specific conditions that generated the observed data. For example, the theoretical covariance between a quantitative behavior and the amount of received reinforcement can generally be obtained from the feedback function of a reinforcement schedule (cf. Borgstede & Luque, 2021). Similarly, the quantitative behavior itself (and, consequently, behavioral change) can be observed over several repeated experimental trials. We can then use these empirically derived parameters to calculate what would have happened in the absence of reinforcement, conditional on the specified model. In other words, assuming that reinforcement can be described as a behavioral selection process and that this process is formally captured by the MLBS, we can calculate the counterfactual behavior by leaving everything in the model unchanged, except for the part that is responsible for selection to occur (cf. Okasha & Otsuka, 2020). Practically, this means to apply a minimal change to the model parameters, such that the covariance term becomes zero (because a zero covariance implies zero selection and hence, no reinforcement effect).

Mathematically, there are several ways to ensure that the covariance term in the CLOE is zero. However, the most plausible candidate for our counterfactual model is certainly that reinforcement is set equal across trials, i.e., we set reinforcement equal to the average reinforcement of all observed trials. For example, if we conduct an experiment with two subsequent trials, we may observe behavior that yields 10 food items in trial one and 20 food items in trial two. The counterfactual model would use the actual feedback functions from the two trials, respectively, to calculate the behavior that would have resulted in the two trials given the individual had continued doing what it was doing, until average reinforcement (i.e., 15 food items) had been received. Conditional on the model, the difference between these two counterfactual values for the behavior

in trial one and two corresponds to the behavior change that we would have observed if there had been no behavioral selection (i.e., no reinforcement). Comparing the change with and without selection, we may calculate an empirical estimate of the causal effect of reinforcement by taking the difference between the counterfactual change and the actually observed change.

The proposed method can thus be summarized by the following steps. First, describing the experimental scenario in terms of behavioral selection using the MLBS. Second, using the model to calculate the amount of behavior change that would have occurred in the absence of selection. Third, subtracting this non-selection component of behavior change from the total behavior change observed between trials to obtain the selection component of behavior change. This latter step corresponds to the causal effect of reinforcement on observed behavior change given the specified model.

The counterfactual methodology provides a new way to infer causal effects of reinforcement from observed behavior change, even if reinforcement is confounded with other causes of behavior change. In the following section, we will fully develop the method by applying it to an actual empirical data set. We will show how empirical estimates of the causal effects of reinforcement can be obtained at the level of individual learning, and how these individual causal effects can be tested for statistical significance using a Monte-Carlo simulation approach.

4 Empirical Application: Behavioral Selection Between Learning Trials

We demonstrate how the counterfactual analysis of selection effects described above can be implemented in an empirical study. Our method is applied to the data from two training trials (first and last days) of a behavioral experiment involving nine naïve pigeons. The focus of the main experiment was to investigate the effects of signals for the availability of food items in the holes of a board on foraging behavior (Anselme et al., in press). Here, we focus on the training trials administered prior to the main experiment, consisting of two conditions only ("no reinforcement" vs. "guaranteed reinforcement" at the beginning of a trial). The question we address is, whether the

pigeons actually learned to distinguish between the reinforcement and the non-reinforcement condition between training trials (i.e., if behavioral selection was effective), or if they primarily relied on non-associative mechanisms like conditional rules to exploit the available food sources. Additionally, we investigate the potential contribution of non-selection sources of behavioral change, such as increased skill and effort.

4.1 Experimental Apparatus and Data Acquisition

4.1.1 Housing

Nine naïve pigeons (4 males and 5 females; age: 5.42±1.15 years) obtained from local breeders were housed in individual cages under a 12h light/dark cycle (lights on at 7:30 am). The pigeons had ad libitum access to water in their home cage but were maintained at 85-90% of their free-feeding body weight for the duration of the experiment, to motivate them for food in the task. They did not receive any additional food besides the supply to the amounts consumed during a training session. All procedures were approved by the Ruhr-University and followed the German guidelines for the care and use of animals in science, and were in accordance with the European Communities Council Directive 86/609/EEC concerning the care and use of animals for experimentation.

4.1.2 Apparatus

The experimental apparatus consisted of a rectangular wooden box with a net on top to prevent the pigeons from flying away. The floor was a horizontally removable brown plate of wood (120 cm length × 70 cm width × 2 cm height), adjusted to the box's size and perforated with holes (1.5 cm diameter and 2 cm depth). The foraging board contained 60 holes organized as 6 rows of 10 holes regularly spaced. The board was covered with a black plastic tape with a cross cut above each hole to create an opening, which allowed the pigeons to access the food items while being unable to visually detect their presence from a distance (Fig. 1a). An entrance compartment (28.7 cm length × 20.4 cm width × 36 cm height) was located in the middle of one long side of the box, which was of the same height. This compartment had two vertical doors on each side, one allowing the experimenter to

362 introduce the pigeon in the compartment and another giving the pigeon access to the board. Specific
363 stimuli (green and red; 21 cm length × 14.5 cm width) were used to signal the consistent presence or
364 absence of food in one area. The two areas were separated by means of a blue-white strip glued on
365 the plastic tape, dividing the board in two equal left and right areas of 30 holes each from the
366 entrance compartment. Each session was recorded with an external camera (GoPro – Hero7), placed
367 above the apparatus.

368 ### 4.1.3 Procedure

369 Before using the board with covered holes (see above), the pigeons had to learn that food items
370 could be found in the board's holes. To this end, the board was initially not covered for 3-4 days,
371 such that the food items could immediately be detected. At this stage, each of the 60 holes was
372 baited with one food item (corn, green pea, yellow pea, or sunflower). No discriminative stimuli and
373 no blue-white strip were used. Following this phase, the covered board was used in association with
374 the discriminative stimuli and the demarcation strip. In each of the 30 holes of one area (left or right,
375 counterbalanced within the same individuals), we positioned one food item and this area was
376 associated with one stimulus (red or green, counterbalanced across individuals) placed on each wall.
377 The 30 holes of the adjacent area remained empty and associated with the other stimulus placed on
378 each wall. The pigeons were trained for four days to learn and inspect the covered holes properly.

379 Practically, a pigeon was placed in the entrance compartment via an external vertical door, which
380 remained closed during each 10-min daily trial. An internal vertical door giving the pigeon access to
381 the board was then immediately removed from top and the compartment remained accessible
382 during the trial—this removal started the trial, during which the pigeon could move on the board,
383 consume grains in the rewarded area, or spend time in the entrance compartment. After a trial, the
384 pigeon was picked up by means of a fishing net from the entrance compartment and the number of
385 food items consumed was manually counted. The board was cleaned and prepared for the next
386 pigeon. After positioning a food item in a hole, the corners of the cut tape covering the hole were

realigned with a tweezer in order to minimize the risk that the food item was detected. Body weight was controlled daily.

4.1.4 Data extraction

Data were collected on manual counting (food items consumed, time spent per area, and number of pecks per area). Determining whether a peck at a hole was successful (food item grasped) or not was mostly impossible from the videos, so that peck is not synonymous with item consumed. A peck simply meant a vertical downshift of the pigeon's head above a hole. We considered a pigeon to be positioned in a given area if its head was in this area—because its body could be in one area and yet pecking in the adjacent one. A crossing between the two areas was not counted if the pigeon came from one area, had a quick look of a few seconds above another area, and then went back to the first one. Sometimes, the pigeon missed a grain (picked it up and lost it), so that it rolled on the board. In the attempt to get it, the pigeon could cross the demarcation line between two areas. The time spent pursuing a missed grain in another area was attributed to the area from which the pigeon came in case it returned to this area after catching the grain. A peck given outside of a hole, even to get a missed grain, was not counted. A pigeon might flap its wings while on the board, often traveling from one area to another. No crossing was counted if the pigeon stopped flapping its wings in the same area as before executing this behavior.

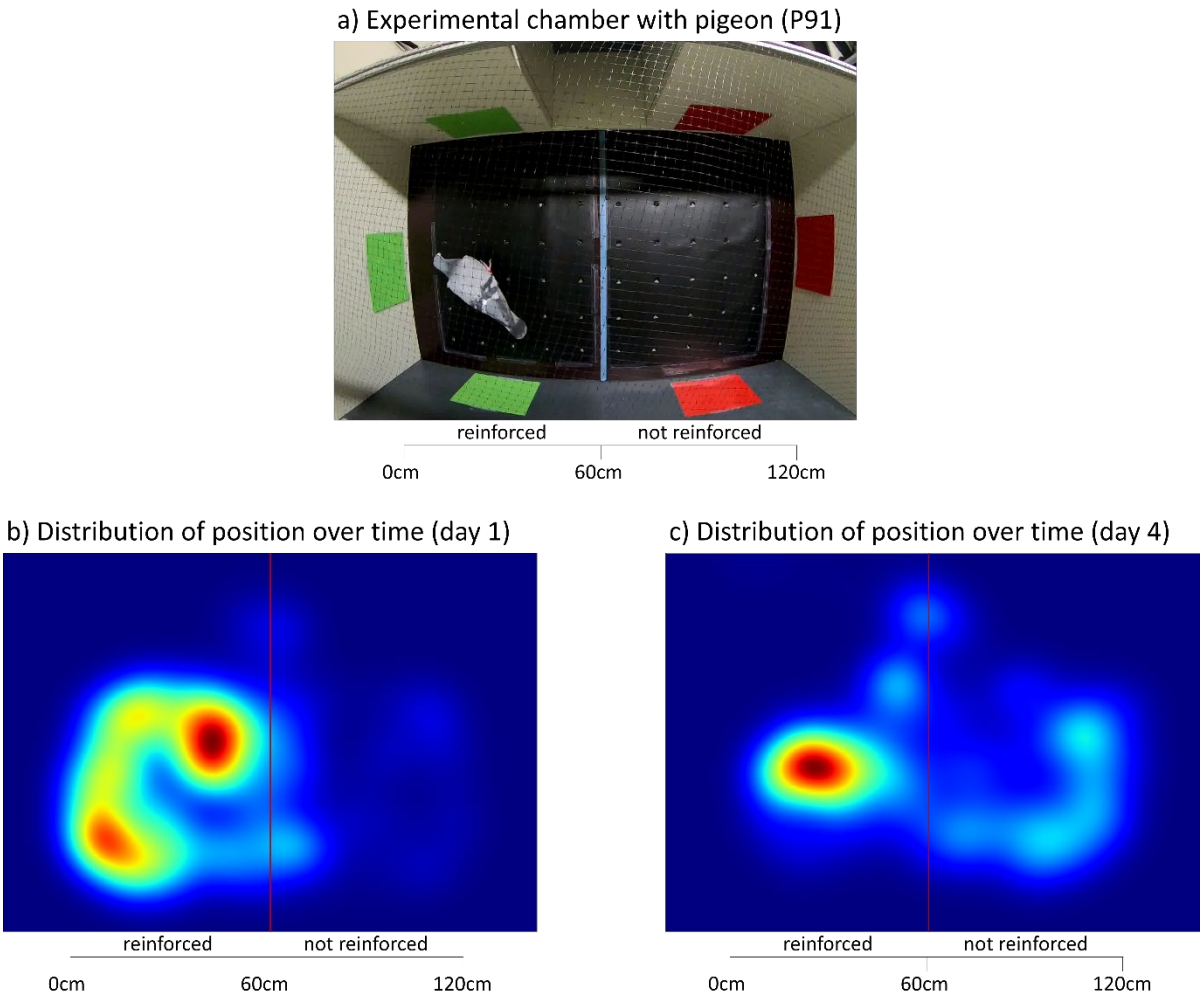

Figure 1: Experimental setup and data collection. (a) Pigeons were put in a 70cmx120cm experimental chamber with 60 covered holes. In the reinforced region, the holes contained food (grains). In the non-reinforced region, the holes were empty. (b, c) Heatmaps show spatial distributions of the pigeon during a 10-minute session for day 1 and day 4, respectively. Duration of visits is indicated by a color gradient ranging from blue (short) over yellow (medium) to red (high). Heatmaps were generated using the ezTrack software (Pennington et al., 2019).

4.2 Statistical Analyses

Pigeons' peck frequency, pecking success and the time spent at the reinforced and non-reinforced region were compared between training trials 1 and 4 using descriptive statistics (mean and standard deviation). Change in average peck frequency and average pecking success were tested for significant deviations from zero ("no change") using one-sided exact permutation tests for paired data to compute p-values for increases in skill and effort, respectively (Edgington & Onghena, 2007). Change in average time spent at the reinforced region was tested for significant deviations from zero ("no change") using two-sided exact permutation tests for paired data.

In order to estimate the causal effect of selection between trials, we parameterized the MLBS under the assumption that reinforcement was absent using the feedback function of the realized schedule of reinforcement and the observed behavior of the pigeons. Since the holes in the experimental chamber were not rebaited during the trials, the probability of finding additional food items decreased over time spent at the reinforced region, depending on how many food items had already been consumed. The rate of decrease in reinforcement probability depends on the pigeon's peck frequency (pecks per time) and the corresponding success rate (probability of retrieving food from a hole, given the hole contains food). Assuming constant peck frequency and constant pecking success within trials, the average feedback function at the reinforced region is:

$$R = N - N(1-\lambda)^t \tag{3.}$$

with R being the expected total amount of reinforcement received until time t, N being the number of food items in the reinforced region at the beginning of the trial, and λ being a free parameter to be calculated from the observed behavior for each experimental trial, separately (see Appendix for mathematical details). Figure 2 depicts the graph of the feedback function for pigeon 91 during the first experimental trial.

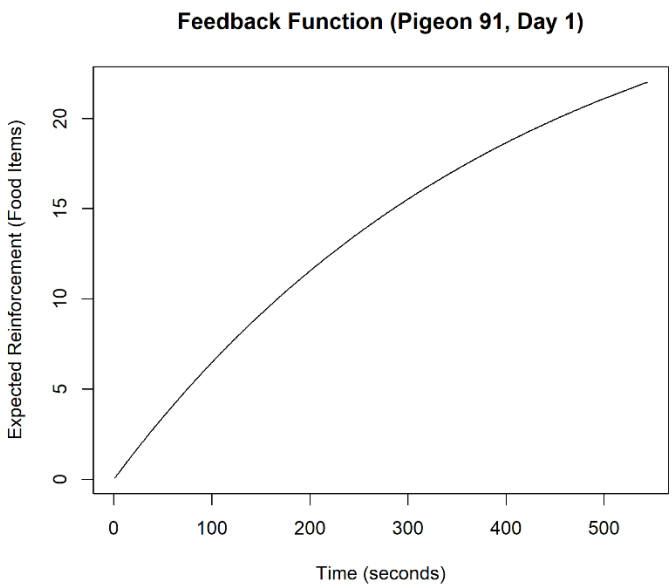

Figure 2: Feedback function in reinforced region given constant effort (pecks/time) and constant skill (consumption/pecks) for Pigeon 91 on the first day of the experiment. Reinforcement at time t is given by $R = N - N(1-\lambda)^t$, with a given

number of 30 food items at the beginning ($N = 30$) and an estimated steepness parameter $\lambda = 0.00242$ (which refers to time measured in seconds).

The individually parameterized feedback functions for the first and the last experimental trials were then used to construct the counterfactual models for each pigeon. Instead of calculating the expected amount of reinforcement from time spent in the reinforced region, the inverse of the feedback function was used to obtain the time it would have taken the pigeon in the respective trial to obtain average reinforcement (i.e., the arithmetic mean between the actually obtained reinforcements during the first and last experimental trials). If the calculated value exceeded the trial length (600 s), the maximum possible value (i.e., 600 s) was used. The difference between these counterfactual values of the time spent at the reinforced region yields the amount of behavior change we would have observed if there had been no behavioral selection. This difference corresponds to the residual term, δ, in the CLOE. Given the actually observed change in time spent in the reinforced region, Δb, and the residual change, δ, from the counterfactual model, behavioral selection, $\Delta_S b$, can be calculated for each pigeon separately using Equation 2.

To test whether the individual selection terms, $\Delta_S b$, are significantly different from zero, the parameterized counterfactual model was simulated for 10000 pairs of trials for each pigeon, separately. In every iteration of the simulation, the time needed to reach average reinforcement in the first and the last training trial was used to calculate the corresponding behavior change. Under the null hypothesis that there is no selection, the distribution of the simulated behavioral change values from the counterfactual model resembles the sampling distribution, from which the actually observed behavior change was drawn. The quantiles of this sampling distribution were used to calculate one-sided p-values for each pigeon (compare Figure 3 and Appendix for technical details). If a p-value yielded a significant result at the 5%-level, the null hypothesis was rejected in favor of the alternative, that the observed change was at least partly caused by selection. The p-values from the individual-level significance tests were further combined multiplicatively using Fisher's method to obtain a meta-analytic result for the significance at a group level (Fisher, 1938). The overall p-value

tests against the null hypothesis that there is no selection effect in the population. All statistical analyses and simulations were conducted using the software R, version 4.0.3 (R Core Team, 2020).

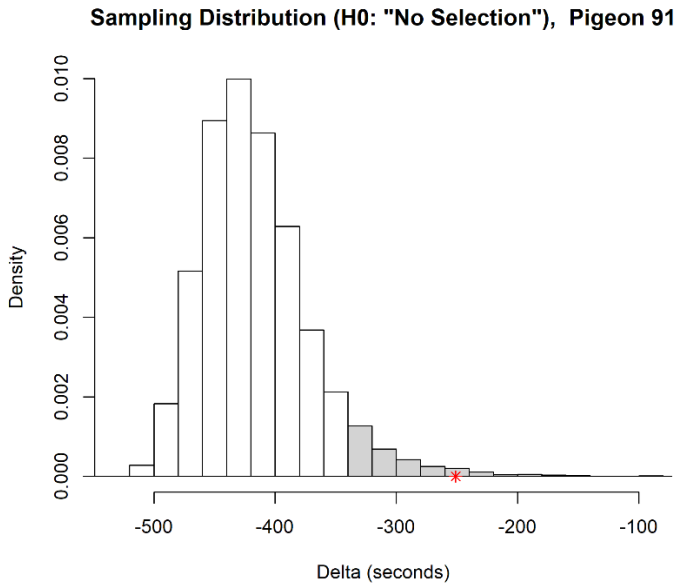

Figure 3: Sampling distribution to test for significant change due to selection between trials (Pigeon 91) obtained from 10000 runs of a Monte-Carlo-Simulation using the counterfactual model (without selection) as the generating model. The expectation of the sampling distribution corresponds to the analytically derived value for $\delta = -426$ s. The gray shading indicates the region of rejection for a one-sided significance test with $H_0: \Delta_S b = 0$. The red star indicates the observed difference in behavior between trials ($\Delta b = -251$ s). Despite this negative change in behavior, the counterfactual model implies that there is in fact a considerable positive selection ($\Delta_S b = 175$ s), which is statistically significant at $\alpha = .05$ ($p = .0087$).

4.3 Results

Table 1 summarizes the pigeons' behavior during the first and the last trial of training, respectively. The main behavioral metric was the time spent at the reinforced region per trial (measured in s). The average time spent at the reinforced region during the first trial was 429 s (*SD* = 99 s), which makes up 72% of the total time (600 s). In other words, pigeons prefer the reinforced region right from the beginning. The average time spent at the reinforced region during the last trial was 305 s (*SD* = 159 s), which makes up only 51% of the total time (600 s). Thus, average time at the reinforced region *decreased* over training trials (p = .027). Average foraging effort, as given by peck frequency while in the reinforced region (pecks per second, measured in s^{-1}), increased from the first trial (M = 0.53 s^{-1}, SD = 0.27 s^{-1}) to the last trial (M = 71 s^{-1}, SD = 0.28 s^{-1}). Hence, average effort increased over trials (p = .015). Average skill, as given by the number of consumed food items per peck (food items per peck,

482 dimensionless number), also increased from the first trial (*M* = 0.07, *SD* = 0.05) to the last trial (*M* =

483 0.13, *SD* = 0.07), indicating an average increase in skill across trials (p = .016). The observed changes

484 in effort and skill are illustrated by Figure 4a and 4b, respectively.

485 *Table 1: Comparison of first and last training trial (day 1 and day 4) with respect to the time spent at the reinforced region,*
486 *the pecks performed per second (peck frequency) and the grains consumed per peck while in the reinforced region (success*
487 *rate).*

	Time at reinforced region[a]		**Peck frequency[b]**		**Success rate[c]**	
	first trial	**last trial**	**first trial**	**last trial**	**first trial**	**last trial**
P17	586 (95 %)	585 (96 %)	0.40	0.15	0.09	0.26
P73	483 (84 %)	382 (63 %)	0.67	0.89	0.04	0.05
P91	544 (89 %)	293 (48 %)	0.84	0.91	0.05	0.11
P96	293 (74 %)	105 (80 %)	0.81	1.11	0.05	0.15
P97	407 (83 %)	187 (76 %)	0.52	0.80	0.05	0.13
P103	413 (72 %)	437 (78 %)	0.11	0.77	0.02	0.06
P118	298 (70 %)	165 (79 %)	0.39	0.63	0.07	0.07
P519	399 (73 %)	410 (74 %)	0.22	0.49	0.2	0.14
P534	441 (79%)	182 (91 %)	0.82	0.64	0.05	0.15
Mean	429	305	0.53	0.71	0.07	0.13
SD	99	159	0.27	0.28	0.05	0.07

488 [a]measured in s, percentage in brackets refers to the total time spent in the reinforced and the nonreinforced region;
489 [b]measured in 1/s; [c] dimensionless number

490

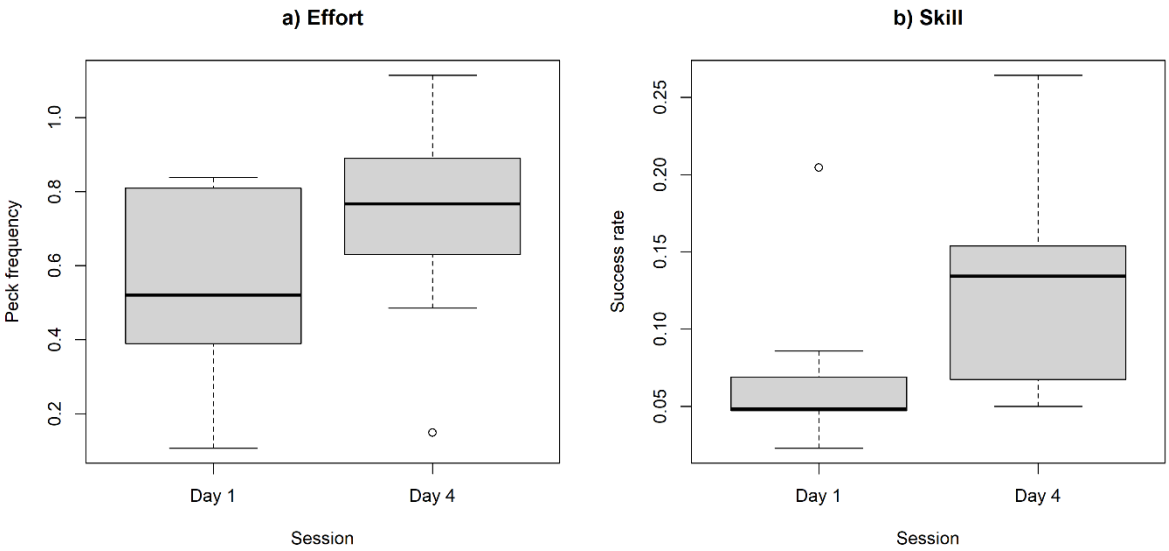

Figure 4: Changes in a) effort (peck frequency measured in pecks per second) and b) skill (success rate measured in consumed food items per peck) from the first day of training to the last day of training.

Table 2 shows the results from the counterfactual analysis. In accordance with the descriptive data, the observed change in time spent at the reinforced region, Δb, was negative for all but two individuals (pigeons 103 and 519), with an average value of -124 s (*SD* = 114 s). The counterfactual model revealed that this negative behavior change is the result of non-selection factors, δ, that were even stronger in the absence of behavioral selection (*M* = -303 s, *SD* = 143 s). Thus, the observed negative change is what remains of these non-selection factors after the effect of behavioral selection has been accounted for. The corresponding selection effects, $\Delta_S b$, were positive for all but one pigeon (pigeon 118). Hence, although the observed change between training trials was generally negative, there was actually positive change due to selection (*M* = 179 s, *SD* = 158 s). The individual-level statistical analyses were significant in four of the nine pigeons, indicating that the data was only conclusive in about half the individuals. However, the meta-analytically combined p-value across all individuals yielded a significant overall selection effect ($p < .0001$). The results from the counterfactual analysis are illustrated in Figure 5, which shows that behavioral selection was generally positive, despite a negative overall change in behavior.

Table 2: Partitioning of total behavioral change in time spent at reinforced region from first to last training day. Total behavioral change (Δb) is expressed as the sum of change due to selection ($\Delta_S b$) and residual change (δ). P-values were calculated using a Monte-Carlo approximation to the sampling distribution under the null hypothesis ($H_0: \Delta_S b = 0$).

	Behavior Change (Δb)	Selection ($\Delta_S b$)	Residual (δ)	Significance[a] (p)
P17	-1	68	-69	.310
P73	-101	136	-237	.240
P91	-251	175	-426	.009*
P96	-188	118	-306	.130
P97	-220	262	-482	.013*
P103	24	442	-418	.0001*
P118	-133	-28	-105	.619
P519	11	391	-380	.0003*
P534	-259	42	-301	.411
Mean[b]	-124	179	-303	< .0001*
SD	114	158	143	–

Δb, $\Delta_S b$ and δ measured in s; [a]approximated p-values obtained from Monte-Carlo simulation; [b]combined p-value calculated using Fisher's method; *statistically significant selection term ($\alpha = .05$)

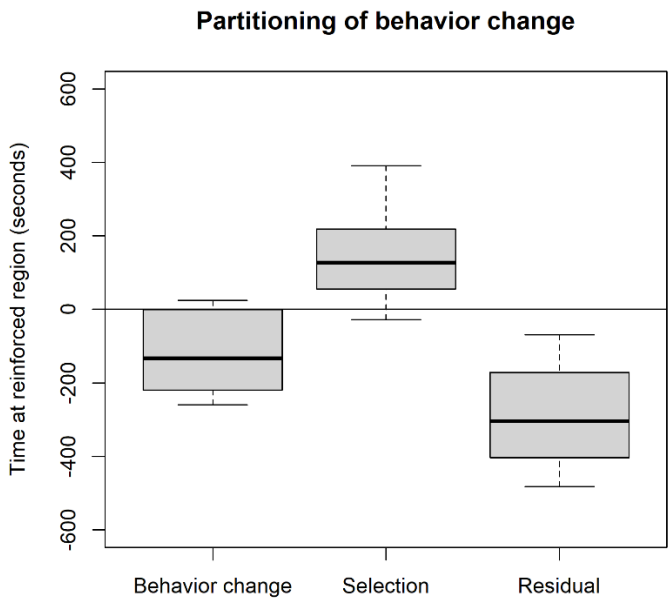

Figure 5: Partitioning of behavior change (time spent at reinforced region) into selection and non-selection (i.e., residual) components using the Covariance Based Law of Effect (CLOE).

5 Discussion

In this paper, we proposed a new method to estimate the causal effects of reinforcement in terms of behavioral selection by means of counterfactual reasoning. The method builds on a formal theory of reinforcement, the Multilevel Model of Behavioral Selection (MLBS), that allows to partition observed changes in average behavior into a behavioral selection component and a non-selection component. Applying the MLBS to empirical data, we showed how the causal effects of reinforcement can be inferred from empirical data, even if reinforcement is masked by other factors. The rationale was to use the MLBS to simulate what would have happened if selection had been absent. The difference between this (counterfactual) behavior change and the observed (actual) behavior change yields an estimate for the causal effect of reinforcement. We further proposed a Monte-Carlo simulation framework to statistically test against the null hypothesis that reinforcement was absent at the level of single individuals. These individual p-values can then be combined using meta-analytic methods.

A first application of the method to the training data from a reinforcement experiment with pigeons revealed that the proposed method is indeed capable to separate selection effects (i.e., behavior change due to reinforcement) from non-selection effects (such as changes due to increasing effort and skill). Although the observed change between training trials was negative (i.e., individuals tended to spend less time at the reinforced region), there was in fact strong evidence for positive behavioral selection. However, these selection effects were masked by significant increases in skill and effort. In other words, the pigeons did not only learn where the food was, but also became more efficient in retrieving the food from the apparatus. Thus, the decrease in time spent in the reinforced region did not result from food acting as a punisher (which would be highly implausible in this context), but from pigeons spending less time foraging due to increased skills. In fact, during the last training trial, several pigeons only engaged in foraging behavior for a rather short time period and spent the rest of the trial with other activities, such as wing flapping or preening. These results indicate that reinforcement, skill and effort all contribute to the observed behavior. Moreover, since all pigeons

543 showed a strong bias towards the reinforced region even in the first trial, it seems likely that the
544 adaptive response of the individuals is at least partly caused by non-associative mechanisms. This
545 conclusion is supported by the qualitative aspects of the pigeons' behavior during the first trial.
546 Pigeons seemed to randomly move through the experimental chamber and occasionally peck, until
547 they encountered food. Subsequently, they would move much less and more slowly, which in turn
548 might have resulted in an increased probability of obtaining more food and an increased amount of
549 time spent at the region where they found food. As mentioned in the introduction, such a behavior
550 might be caused by a local decision rule ("If you find food, move less!"), rather than reinforcement.

551 The existence of such a rule is compatible with the marginal value theorem from optimal foraging
552 theory (Charnov, 1976). It predicts that a patch will be abandoned when its resources deplete to the
553 point that the next expected reward falls below the average reward in the patch. It also predicts that
554 a longer transit time between two patches will maintain the forager longer in the depleting patch.
555 Here, the transit time between the two adjacent areas was short but one of them contained no food
556 at all, giving a signal to the pigeons that there was no patch of interest around. This is likely to have
557 kept the pigeons in the reinforced area. A similar propensity to spend time, peck, and visit the food-
558 richest area compared to poorer areas was also shown (Anselme et al., in press). Furthermore, in the
559 presence of a single patch type, pigeons visit adjacent holes more often when food density is higher
560 at the beginning of a trial (Wittek et al., 2022). The rule "If you find food, move less!" would explain
561 these results, even if it was hardwired in the pigeon's brain and thus independent of individual
562 experiences. However, the rule itself does not tell us how individual pigeons convert it into action.
563 Reinforcement could be a second-order process based on this rule, capable of influencing behavior at
564 an individual level. Indeed, our findings suggest considerable inter-individual variation in both the
565 selection and non-selection effects, which might be related to distinct "temperaments" among the
566 individuals.

567 We present the first empirical application of the MLBS in the context of a reinforcement experiment.
568 The results show that the MLBS provides a feasible theoretical foundation for the experimental

analysis of behavior and thereby demonstrate the strength of the selectionist approach to reinforcement. Future work may extend the method to further disentangle non-selection effects such as skill and effort, alongside with other potential sources of behavioral change.

The proposed selection framework acknowledges that reinforcement cannot be isolated as an independent variable. The resulting methodology allows to infer causal effects of reinforcement even if behavior is in a non-equilibrium state. Shifting the focus away from steady-state analysis (such as the application of the generalized matching law), the proposed method specifically applies to behavior that is currently changing, such as the adaptive dynamics during operant training. In general, data from training trials are rather messy due to various uncontrollable factors affecting behavior simultaneously. We think that the counterfactual analysis of behavioral selection provides the necessary tools to deal with these data. However, it should not be concluded that non-selection effects necessarily operate against predictors of reinforcement such as food. Among them, indirect predictors of reinforcement could exist, such as information about the consistency of cue-food pairings. For example, in a free-choice task, pigeons often prefer an option in which a consistent cue-food pairing occurs rarely (e.g., red probability = 20% but if red $\rightarrow$ 100% chance of one food item) to another option in which more food is delivered but the cue-food pairing is inconsistent (e.g., blue and yellow probability = 50% and whether blue or yellow $\rightarrow$ 50% chance of three food items) (Anselme, 2022; Fortes et al., 2016; McDevitt et al., 2016; Zentall, 2016). Parameters that control information seeking might be relevant predictors of reinforcement even if their influence may be less direct and possibly independent from food (e.g., Inglis et al., 1997).

The general methodology of model-based counterfactual analysis can be applied to many other experimental setups that involve behavioral change over time – probably, there are already thousands of unused training data sets only awaiting to be analyzed. The partitioning of behavior change into selection and non-selection components may equally be applied in naturalistic settings, such as actual foraging behavior. Moreover, the method provides a promising framework to analyze the complexities of human behavior, such as the relation between reinforcement and competence

gain in students' learning behavior or the interaction between reinforcement and verbal instruction in the formation of rule-governed behavior. We hope that this article contributes to the foundations of behavioral selection as a general theory of behavior, and encourages other researchers to put the behavioral selection perspective into practice.

6 Author Contributions

XX developed the modeling framework and the statistical methodology, analyzed the data and wrote the original draft. XY planned and conducted the experiments, collected the data and contributed to the theoretical background and the interpretation of the results. Both authors contributed to the final version of the manuscript.

7 Competing Interests

The authors declare no competing interests.

8 Data and Code Availability

The original data and R-code used in this study will be made publicly available on publication.

9 Funding

This research was supported by the Deutsche Forschungsgemeinschaft through An1067/3-1 (PA).

10 Appendix

10.1 Derivation of Feedback Function

At each time step, there are $N(t)$ food items present in the reinforced region. Given that a single food item is consumed with probability λ, the expected value of reinforcement at each time step is $N(t)\lambda$. We can thus express the expected change in reinforcement from one time step (t) to the next time step ($t+1$) by the following difference equation:

$$\Delta R(t) = N(t)\lambda$$

The corresponding recursion equation is:

$$R(t+1) = R(t) + N(t)\lambda$$

Since $R(t) = N - N(t)$, by substitution we obtain:

$$R(t+1) = N - N(t) + N(t)\lambda = N - N(t)(1-\lambda)$$

The general solution of this dynamic equation is:

$$R(t) = N - N(1-\lambda)^t$$

This corresponds to the feedback function of expected reinforcement given constant peck rate and constant pecking success.

10.2 Calculation of individual p-values

Assuming that time steps are small enough, such that an individual can never consume more than one food item per time step, the probability of being reinforced in a given time step equals the probability of at least one success, which is given by $1 - (1-\lambda)^{N(t)}$.

We can thus simulate the pigeon's behavior as a series of random reinforcement events with a reinforcement probability of $1 - (1-\lambda)^{N(t)}$ at each time step. The parameter λ is estimated from the observed data using the feedback function. $N(t)$ is calculated as the result of the previous reinforcement events, i.e., starting at the total number of available food items, $N(t)$ is reduced by 1 in every time step that yields reinforcement. The simulation is run for the first and last training trial (with λ being estimated from the observed behavior in the corresponding trial), until average reinforcement (i.e., the arithmetic mean of the observed reinforcement over both trials) or the maximum possible time (i.e., the length of one experimental trial) are reached.

The simulation was repeated 10000 times to approximate the sampling distribution, from which the actually observed behavior has been drawn. The quantiles of this sampling distribution yield a critical

value for the region of rejection, as well as the approximate p-value of the observed data, given the data was generated by the specified model.

11 References

Anselme, P. (2022). The optimality of "suboptimal" choice: A psycho-evolutionary perspective. In M. Krause, K. L. Hollis, & M. R. Papini (Eds.), *Evolution of Learning and Memory Mechanisms* (pp. 193–209). Cambridge University Press.

Anselme, P., Dreher, T., & Güntürkün, O. (2018). Pigeons consistently prefer easy over harder access to food: No reversal after direct dopaminergic stimulation. *Behavioral Neuroscience*, *132*(4), 293–301. https://doi.org/10.1037/bne0000249

Anselme, P., Wittek, N., Oeksuez, F., & Güntürkün, O. (in press). Overmatching under food uncertainty in foraging pigeons. *Behavioural Processes*.

Baum, W. M. (2012). Rethinking reinforcement: Allocation, induction, and contingency. *Journal of the Experimental Analysis of Behavior*, *97*(1), 101–124. https://doi.org/10.1901/jeab.2012.97-101

Baum, W. M. (2017). Selection by consequences, behavioral evolution, and the price equation. *Journal of the Experimental Analysis of Behavior*, *107*(3), 321–342. https://doi.org/10.1002/jeab.256

Berlyne, D. E. (1957). Uncertainty and conflict: A point of contact between information-theory and behavior-theory concepts. *Psychological Review*, *64, Part 1*(6), 329–339. https://doi.org/10.1037/h0041135

Berridge, K. C., & Robinson, T. E. (2016). Liking, wanting, and the incentive-sensitization theory of addiction. *The American Psychologist*, *71*(8), 670–679. https://doi.org/10.1037/amp0000059

Borgstede, M. (2020). An evolutionary model of reinforcer value. *Behavioural Processes*, 104109. https://doi.org/10.1016/j.beproc.2020.104109

Borgstede, M. (2021). Why do individuals seek information? A selectionist perspective. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2021.684544

Borgstede, M., & Eggert, F. (in press). Squaring the Circle: From Latent Variables to Theory Based Measurement. *Theory & Psychology*.

Borgstede, M., & Eggert, F. (2021). The formal foundation of an evolutionary theory of reinforcement. *Behavioural Processes*, *186*, 104370. https://doi.org/10.1016/j.beproc.2021.104370

Borgstede, M., & Luque, V. J. (2021). The covariance based law of effect: A fundamental principle of behavior. *Behavior and Philosophy*, *49*, 63–81. https://behavior.org/wp-content/uploads/2022/01/BP-v49-Borgstede2.pdf

Broadbent, D. E. (1961). *Behaviour*. Methuen.

Campbell, D. T. (1956). Adaptive behavior from random response. *Behavioral Science*, *1*(2), 105–110. https://doi.org/10.1002/bs.3830010204

Charnov, E. L. (1976). Optimal foraging, the marginal value theorem. *Theoretical Population Biology*, *9*(2), 129–136. https://doi.org/10.1016/0040-5809(76)90040-X

Davison, M., & McCarthy, D. (2016). *The Matching Law*. Routledge. https://doi.org/10.4324/9781315638911

Dugard, P., File, P., & Todman, J. B. (2012). *Single-case and small-n experimental designs: A practical guide to randomization tests* (2. ed.). Routledge.

Edgington, E., & Onghena, P. (2007). *Randomization Tests, Fourth Edition* (4th ed.). *Statistics*. CRC Press.

Eilam, D., Zor, R., Szechtman, H., & Hermesh, H. (2006). Rituals, stereotypy and compulsive behavior in animals and humans. *Neuroscience and Biobehavioral Reviews*, *30*(4), 456–471. https://doi.org/10.1016/j.neubiorev.2005.08.003

Ferster, C. B., & Skinner, B. F. (1957). *Schedules of reinforcement*. Appleton-Century-Crofts. https://doi.org/10.1037/10627-000

Fisher, R. A. (1938). *Statistical methods for research workers* (7. ed., rev. and enl). *Biological monographs and manuals*. Oliver & Boyd.

Fortes, I., Vasconcelos, M., & Machado, A. (2016). Testing the boundaries of "paradoxical" predictions: Pigeons do disregard bad news. *Journal of Experimental Psychology. Animal Learning and Cognition*, *42*(4), 336–346. https://doi.org/10.1037/xan0000114

Gilbert, R. M. (1970). Psychology and biology. *Canadian Psychologist/Psychologie Canadienne*, *11*(3), 221–238. https://doi.org/10.1037/h0082574

Hendry, D. P. (1965). *Reinforcing Value of Information: NASA Technical Report No. 65-1.* University of Maryland.

Herrnstein, R. J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior*, *4*, 267–272. https://doi.org/10.1901/jeab.1961.4-267

Herrnstein, R. J. (1964). Will. *Proceedings of the American Philosophical Society,*, *108*(6), 455–458.

Inglis, Forkman, & Lazarus (1997). Free food or earned food? A review and fuzzy model of contrafreeloading. *Animal Behaviour*, *53*(6), 1171–1191. https://doi.org/10.1006/anbe.1996.0320

Kamin, L. J. (1969). Predictability, surprise, attention and conditioning. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior* (pp. 279–296).

Lemos, J. C., Wanat, M. J., Smith, J. S., Reyes, B. A. S., Hollon, N. G., van Bockstaele, E. J., Chavkin, C., & Phillips, P. E. M. (2012). Severe stress switches CRF action in the nucleus accumbens from appetitive to aversive. *Nature*, *490*(7420), 402–406. https://doi.org/10.1038/nature11436

Lewis, D. (1973). Causation. *The Journal of Philosophy*, *70*(17), 556. https://doi.org/10.2307/2025310

Lewis, D. (2000). Causation as Influence. *The Journal of Philosophy*, *97*(4), 182. https://doi.org/10.2307/2678389

Luque, V. J. (2017). One equation to rule them all: a philosophical analysis of the Price equation. *Biology & Philosophy*, *32*(1), 97–125. https://doi.org/10.1007/s10539-016-9538-y

Luque, V. J., & Baravalle, L. (2021). The mirror of physics: on how the Price equation can unify evolutionary biology. *Synthese*, *199*, 12439–12462. https://doi.org/10.1007/s11229-021-03339-6

McDevitt, M. A., Dunn, R. M., Spetch, M. L., & Ludvig, E. A. (2016). When good news leads to bad choices. *Journal of the Experimental Analysis of Behavior*, *105*(1), 23–40. https://doi.org/10.1002/jeab.192

McNamara, J. M., & Houston, A. I. (2009). Integrating function and mechanism. *Trends in Ecology & Evolution*, *24*(12), 670–675. https://doi.org/10.1016/j.tree.2009.05.011

Menzies, P., & Price, H. (1993). Causation as a Secondary Quality. *The British Journal for the Philosophy of Science*, *44*(2), 187–203. https://doi.org/10.1093/bjps/44.2.187

Okasha, S., & Otsuka, J. (2020). The Price equation and the causal analysis of evolutionary change. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *375*(1797), 20190365. https://doi.org/10.1098/rstb.2019.0365

Palmer, D. C., & Donahoe, J. W. (1992). Essentialism and selectionism in cognitive science and behavior analysis. *The American Psychologist*, *47*(11), 1344–1358. https://doi.org/10.1037/0003-066X.47.11.1344

Pearl, J. (2013). *Causality: Models, reasoning, and inference* (Second edition, reprinted with corrections). Cambridge Univ. Press.

Pennington, Z. T., Dong, Z., Feng, Y., Vetere, L. M., Page-Harley, L., Shuman, T., & Cai, D. J. (2019). Eztrack: An open-source video analysis pipeline for the investigation of animal behavior. *Scientific Reports*, *9*(1), 19979. https://doi.org/10.1038/s41598-019-56408-9

Premack, D., & Premack, A. J. (1963). Increased eating in rats deprived of running. *Journal of the Experimental Analysis of Behavior*, *6*, 209–212. https://doi.org/10.1901/jeab.1963.6-209

Price, G. R. (1970). Selection and Covariance. *Nature*, *227*(5257), 520–521. https://doi.org/10.1038/227520a0

Price, G. R. (1972). Extension of covariance selection mathematics. *Annals of Human Genetics*, *35*(4), 485–490. https://doi.org/10.1111/j.1469-1809.1957.tb01874.x

Pringle, J. (1951). On the Parallel Between Learning and Evolution. *Behaviour*, *3*(1), 174–214. https://doi.org/10.1163/156853951X00269

R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. https://www.R-project.org/

Skinner, B. F. (1965). *Science and Human Behavior*. Free Press. https://ebookcentral.proquest.com/lib/gbv/detail.action?docID=4934291

Skinner, B. F. (1966). The phylogeny and ontogeny of behavior. Contingencies of reinforcement throw light on contingencies of survival in the evolution of behavior. *Science (New York, N.Y.)*, *153*(3741), 1205–1213. https://doi.org/10.1126/science.153.3741.1205

Skinner, B. F. (1981). Selection by consequences. *Science (New York, N.Y.)*, *213*(4507), 501–504.

Staddon, J. E. R. (2020). The Role of Theory in Behavior Analysis: A Response to Unfinished Business, Travis Thompson's Review of Staddon's New Behaviorism (2nd edition). *The Psychological Record*, 1–7. https://doi.org/10.1007/s40732-020-00409-y

Strand, P. S., Robinson, M. J. F., Fiedler, K. R., Learn, R., & Anselme, P. (2021). Quantifying the instrumental and noninstrumental underpinnings of Pavlovian responding with the Price equation. *Psychonomic Bulletin & Review.* Advance online publication. https://doi.org/10.3758/s13423-021-02047-z

Thorndike, E. L. (1900). The associative processes in animals. *Biological Lectures from the Marine Biological Laboratory of Woods Holl, 1899*, 69–91.

Timberlake, W., & Allison, J. (1974). Response deprivation: An empirical approach to instrumental performance. *Psychological Review*, *81*(2), 146–164. https://doi.org/10.1037/h0036101

Wang, M. Z., & Hayden, B. Y. (2019). Monkeys are curious about counterfactual outcomes. *Cognition*, *189*, 1–10. https://doi.org/10.1016/j.cognition.2019.03.009

Wittek, N., Oeksuez, F., Güntürkün, O., & Anselme, P. (2022). More opportunities to peck for identical food availability increases foraging efficiency in pigeons. *Behaviour*, 1–24. https://doi.org/10.1163/1568539X-bja10173

Zentall, T. R. (2016). Resolving the paradox of suboptimal choice. *Journal of Experimental Psychology. Animal Learning and Cognition*, *42*(1), 1–14. https://doi.org/10.1037/xan0000085